Inference
A Contribution to the collection "Stochastic Geometry: Highlights, Interactions and New Perspectives"
Møller, Jesper

Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
? You may not further distribute the material or use it for any profit-making activity or commercial gain
? You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: September 07, 2020
Inference
—a contribution to the collection “Stochastic Geometry: Highlights, Interactions and New Perspectives”

by

Jesper Møller

R-2007-22 September 2007
0.1 Inference

(This text written by Jesper Møller, Aalborg University, is submitted for the collection ‘Stochastic Geometry: Highlights, Interactions and New Perspectives’, edited by Wilfrid S. Kendall and Ilya Molchanov, to be published by Clarendon Press, Oxford, and planned to appear as Section 4.1 with the title ‘Inference’.)

This contribution concerns statistical inference for parametric models used in stochastic geometry and based on quick and simple simulation-free procedures as well as more comprehensive methods using Markov chain Monte Carlo (MCMC) simulations. Due to space limitations the focus is on spatial point processes.

0.1.1 Spatial point processes and other random closed set models

Recall that a spatial point process $X$ considered as a random closed set in $\mathbb{R}^d$ is nothing but a random locally finite subset of $\mathbb{R}^d$. This means that the number of points $n(X \cap B)$ falling in an arbitrary bounded Borel set $B \subset \mathbb{R}^d$ is a finite random variable. This extends to a marked point process where to each point $x_i \in X$ there is associated a mark, that is a random variable $K_i$ defined on some measurable space $M$ (for details, see e.g. Stoyan et al. (1995) or Daley and Vere-Jones (2003)).

Most theory and methodology for inference in stochastic geometry concern spatial point processes and to some extend marked point processes, cf. the monographs Ripley (1981, 1988), Cressie (1993), Stoyan et al. (1995), Stoyan and Stoyan (1995), Van Lieshout (2000), Diggle (2003), Møller and Waagepetersen (2003), and Baddeley et al. (2006). A particular important case is a germ-grain model (Hanisch, 1981), where the $x_i$ are called germs and the $K_i$ primary grains, the mark space $M = K$ is the set of compact subsets of $\mathbb{R}^d$, and the object of interest is the random closed $U$ set given by the union of the translated primary grains $x_i + K_i = \{x_i + x : x \in K_i\}$. The most studied case is the Boolean model (Hall, 1988; Molchanov, 1997), i.e. when the germs form a Poisson process and the primary grains are mutually independent, identically distributed, and independent of the germs. Extensions to models with interacting grains have been studied in Kendall et al. (1999) and Møller and Helisova (2007).

0.1.2 Outline and some notation

In the sequel we mostly confine attention to planar point processes, but many concepts, methods, and results easily extend to $\mathbb{R}^d$ or a more general metric space, including many marked point process models. (For example, when discussing the Norwegian spruces in Figure 0.6, this may be viewed as a marked point process of discs.) We concentrate on the two most important classes of spatial point process models, namely Poisson/Cox processes in Section 0.1.3 and Gibbs (or Markov) point processes in Section 0.1.4. We illustrate the statistical methodology with various application examples, where many are examples of inhomogeneous point patterns. We discuss the current state of simulation-based maximum likelihood inference as well as Bayesian inference, where fast computers and advances in computational statistics, particularly MCMC meth-
ods, have had a major impact on the development of statistics for spatial point processes. The MCMC algorithms are treated in some detail; for a general introduction to MCMC and for spatial point processes in particular, see Møller and Waagepetersen (2003) and the references therein. We also discuss more classical methods based on summary statistics, and more recent simulation-free inference procedures based on composite likelihoods, minimum contrast estimation, and pseudo likelihood. Often the R package spatstat (Baddeley and Turner 2005, 2006) is used; in other cases own software in R and C has been developed.

Throughout this contribution we use the following notation. Unless otherwise stated, \( X \) is a planar spatial point process and \( B \) is the class of Borel sets in \( \mathbb{R}^2 \). We let \( W \in B \) denote a bounded observation window of area \( |W| > 0 \), and assume that a realization \( X \cap W = \mathbf{x} \) is observed (most often only one such realization is observed). Here the number of points, denoted \( n(\mathbf{x}) \), is finite. Moreover, \( I[\cdot] \) (or \( \mathbb{I}[\cdot] \)) is an indicator function.

0.1.3 Inference for Poisson and Cox process models

Poisson processes (Kingman, 1993) are models for point patterns with complete spatial independence, while Cox processes (Cox, 1955) typically model a positive dependence between the points. Both kinds of processes may model inhomogeneity (Diggle, 2003; Møller and Waagepetersen 2003, 2007).

As an illustrative example of an inhomogeneous point pattern we refer to the rain forest trees in Figure 0.1. These data have previously been analyzed in Waagepetersen (2007a) and Møller and Waagepetersen (2007), and they are just a small part of a much larger data set comprising hundreds of thousands of trees belonging to hundreds of species (Hubbell and Foster, 1983; Condit et al., 1996; Condit, 1998).

![Image of locations of trees](image_url)

Fig. 0.1. Locations of 3605 Beilschmiedia pendula Lauraceae trees observed within a 500 × 1000 m region at Barro Colorado Island.

0.1.3.1 Definition and fundamental properties of Poisson and Cox processes

The most fundamental spatial point process model is the Poisson process in \( \mathbb{R}^2 \).
The process is defined in terms of a locally finite and diffuse measure $\mu$ on $\mathcal{B}$. These assumptions on $\mu$ allow us to view the Poisson process as a locally finite random closed set $X \subseteq \mathbb{R}^2$; for more general definitions, relaxing the assumptions on $\mu$, see e.g. Kingman (1993). Below we recall the definition of the Poisson process, and discuss why extensions to Cox processes are needed.

In stochastic geometry terms, the Poisson process is simplest characterized by its avoidance probability

$$
P(X \cap B = \emptyset) = \exp(-\mu(B)), \quad B \in \mathcal{B} \tag{0.1}
$$

showing that the process is uniquely defined by $\mu$. A more constructive description is that

(i) $N(B) = n(X \cap B)$, the number of points falling within any bounded $B \in \mathcal{B}$, is Poisson distributed with mean $\mu(B)$, and

(ii) conditional on $N(B)$, the points in $X \cap B$ are mutually independent and identically distributed with a distribution given by the normalized restriction of $\mu$ to $B$.

Clearly (i) implies (0.1) and $\mu$ is the intensity measure of the Poisson process, and (ii) implies that the locations of the points in $X \cap B$ are independent of $N(B)$. It is rather straightforwardly verified that (i)-(ii) is equivalent to (i) and (iii), where (iii) is the independent scattering property:

(iii) $N(A)$ and $N(B)$ are independent if $A, B \in \mathcal{B}$ are disjoint.

Considering an arbitrary countable partition of $\mathbb{R}^2$ into bounded Borel sets and applying (i)-(iii) is one way of constructing and thereby verifying the existence of the Poisson process, see e.g. Møller and Waagepetersen (2003).

Throughout this contribution, as in most statistical applications, $\mu$ is assumed to be absolutely continuous with respect to Lebesgue measure, i.e. $\mu(B) = \int_B \rho(x) \, dx$, where $\rho$ is the intensity function. If $\rho(x) = c$ is constant for all $x$ in a set $B \in \mathcal{B}$, then the Poisson process is said to be homogeneous on $B$ with intensity $c$. Note that stationarity of the Poisson process (i.e. that its distribution is invariant under translations in $\mathbb{R}^2$ of the point process) is equivalent to the process being homogeneous on $\mathbb{R}^2$.

Although various choices of inhomogeneous intensity functions may generate different kinds of aggregated point patterns, the Poisson process is usually considered to be a too simplistic model for statistical applications because of the strong independence properties described in (ii)-(iii) above. A natural extension giving rise to a model for aggregated point patterns with more dependence structure is given by a doubly stochastic construction, where $\Lambda(x)$ is a random locally integrable function so that $X$ conditional on a realization $\Lambda(x) = \lambda(x)$, $x \in \mathbb{R}^2$, is a Poisson process with intensity function $\lambda(x)$. Then $X$ is said to be a Cox process (Cox, 1955) driven by $\Lambda$ (or driven by the random measure given by $M(B) = \int_B \Lambda(x) \, dx$). The simplest example is a mixed Poisson process, i.e. when $\Lambda(x) = \Lambda(o)$ does not depend on $x$ where $\Lambda(o)$ is a non-negative random
variable, see Grandell (1997). In practice more flexible models are needed, as discussed in the following section.

0.1.3.2 Modelling intensity

In order to model inhomogeneity, spatial heterogeneity or aggregation it is important to account for possible effects of covariates (also called explanatory variables), which we view as a non-random \((p + 1)\)-dimensional real function \(z(x) = (z_0(x), \ldots, z_p(x)), x \in \mathbb{R}^2\), and which we assume is known at least within the observation window \(W\). In practice \(z(x)\) may only be partially observed on a grid of points, and hence some interpolation technique may be needed (Rathbun, 1996; Rathbun et al., 2007; Waagepetersen, 2007c). For instance, Figure 0.2 shows two kinds of covariates \(z_1\) and \(z_2\) for the rain forest trees in Figure 0.1, taking \(z_0 \equiv 1\) as discussed below and letting \(z_1\) and \(z_2\) be constant on each of \(100 \times 200\) squares of size \(5 \times 5\) m.

![Figure 0.2](image)

**Fig. 0.2.** Rain forest trees: the covariates \(z_1\) (altitude; left panel) and \(z_2\) (norm of altitude gradient; right panel) are recorded on a 5 by 5 m grid.

Poisson processes with log-linear intensity: The mathematical most convenient model for a Poisson process is given by a log-linear intensity function (Cox, 1972)

\[
\rho(x) = \exp(\beta \cdot z(x))
\]

(0.2)

where \(\beta = (\beta_0, \ldots, \beta_p)\) is a real \((p + 1)\)-dimensional parameter and \(\cdot\) denotes the usual inner product. We refer to \(\beta_0, \ldots, \beta_p\) as regression parameters. In the sequel, as in most applications, we assume that \(z_0 \equiv 1\) and call \(\beta_0\) the intercept. Thus the process is in general inhomogeneous if \(p > 0\).

Log-Gaussian Cox processes: Let \(\Phi = \{\Phi(x) : x \in \mathbb{R}^2\}\) denote a stationary Gaussian process (Adler, 1981). The log-linear Poisson process model extends naturally to a Cox process driven by

\[
\Lambda(x) = \exp(\beta \cdot z(x) + \Phi(x)).
\]

(0.3)

Such log-Gaussian Cox process models (Coles and Jones, 1991; Møller et al., 1998) provide a lot of flexibility corresponding to different types of covariance functions \(c\) where by stationarity of \(\Phi\), \(c(x - y) = \Cov(\Phi(x), \Phi(y))\). Henceforth,
since we can always absorb the mean of $\Phi$ into the intercept $\beta_0$, we assume without loss of generality that $E\Phi(x) = -c(0)/2$ or equivalently that $E\exp(\Phi(x)) = 1$. Thereby the log-Gaussian Cox process also has intensity function (0.2), cf. Section 0.1.3.5.

**Shot-noise Cox processes:** Another very flexible class of models is the shot-noise Cox processes (Wolpert and Ickstadt, 1998; Brix, 1999; Møller, 2003). As an example of this, let

$$\Lambda(x) = \exp(\beta \cdot z(x)) \frac{1}{\omega \sigma^2} \sum_{y \in Y} k((x - y)/\sigma)$$

(0.4)

where $Y$ is a stationary Poisson process with intensity $\omega > 0$, $k(\cdot)$ is a density function with respect to Lebesgue measure, and $\sigma > 0$ is a scaling parameter. This form of the random intensity function can be much generalized, replacing $Y$ with another kind of model and the kernel $k(x - y)$ by another density $k(\cdot; y)$ which may be multiplied by a positive random variable $\gamma_y$, leading to (generalized) shot-noise Cox processes (Møller and Torrisi, 2005). In analogy with the assumption made above for log Gaussian Cox processes, the random part following after the exponential term in (0.4) is a stationary stochastic process with mean one, so the shot-noise Cox process also has intensity function (0.2), cf. Section 0.1.3.5. Often a bivariate normal kernel $k(x) = \exp(-\|x\|^2/2)/(2\pi)$ is used, in which case we refer to (0.4) as a modified Thomas process (Thomas, 1949).

**Simulation** Due to the complexity of spatial point process models, simulations are often needed when fitting a model and studying the properties of various statistics such as parameter estimates and summary statistics. For a Poisson process $X$ with intensity function $\rho$, it is useful to notice that an independent thinning $\{x \in X : U(x) \leq p(x)\}$, where $p : \mathbb{R}^2 \to [0,1]$ is a Borel function and the $U(x)$ are mutually independent uniform random variables on $[0,1]$ which are independent of $Y$, is a Poisson process with intensity function $p\rho$.

Simulation of a stationary Poisson process $X$ restricted to a bounded region $B$ is straightforward, using (i)-(ii) in Section 0.1.3.1. If $B$ is of irregular shape, we simulate $X$ on a larger region $D \supset B$ such as a rectangle or a disc, and return $X \cap B$ (corresponding to an independent thinning with retention probabilities $p(x) = 1[x \in B]$).

Usually the intensity function $\rho(x)$ of an inhomogeneous Poisson process $X$ is bounded by some constant $c$ for all $x \in B$. Then we can first simulate a homogeneous Poisson process on $B$ with intensity $c$, and second make an independent thinning of this with retention probabilities $p(x) = \rho(x)/c$.

Thus the essential problem of simulating a Cox process on $B$ is how to simulate its random intensity function $\Lambda(x)$ for $x \in B$. For the log-Gaussian Cox process driven by (0.3), there exist many ways of simulating the Gaussian process (Schlater, 1999; Lantuejoul, 2002; Møller and Waagepetersen, 2003). For
the shot-noise Cox process $X$ driven by (0.4), it is useful to notice that this is a Poisson cluster process obtained as follows. Conditional on $Y$ in (0.4), let $X_y$ for all $y \in Y$ be mutually independent Poisson processes, where $X_y$ has intensity function $\rho_y(x) = \exp(\beta \cdot z(x)) k((x - y)/\sigma)/\omega \sigma^2$. Then we can take $X = \bigcup_{y \in Y} X_y$, that is, the superposition of ‘clusters’ $X_y$ with ‘centres’ $y \in Y$.

We may easily simulate each cluster, but how do we simulate those centre points generating clusters with points in $B$? In fact such centre points form a Poisson process with intensity function $\rho(y) = \omega P(X_y \cap B \neq \emptyset)$, and the paper by Brix and Kendall (2002) shows how we can simulate this process as well as the non-empty sets $X_y \cap B$. This paper provides indeed a neat example of a perfect simulation algorithm, but it is easier to use an algorithm for approximate simulation, simulating only those centres which appear within a bounded region $D \supset B$ such that $P(X_y \cap B \neq \emptyset)$ is very small if $y \not\in D$ (Møller, 2003; Møller and Waagepetersen, 2003).

0.1.3.4 Maximum likelihood Consider a parametric point process model specified by an unknown parameter $\theta$, e.g. $\theta = \beta$ in case of (0.2); many other examples appear in the sequel. We assume that the point process $X$ restricted to $W$ has a density $f_{\theta}$ with respect to a stationary Poisson process $Y$ with intensity one, i.e. $P(X \cap W \in F) = E(f_{\theta}(Y \cap W)|[Y \cap W \in F])$ for all events $F$. Given the data $X \cap W = x$, the likelihood function $L(\theta; x)$ is any function of $\theta$ which is proportional to $f_{\theta}(x)$, and the maximum likelihood estimate (MLE) is the value of $\theta$ which maximizes the likelihood function (provided such a value exists and it is unique). Thus when we specify the likelihood function we may possibly (and without mentioning) exclude a multiplicative term of the density, if this term depends only on $x$ (and not on $\theta$). Obviously, the MLE does not depend on such a multiplicative term.

The log-likelihood function of the Poisson process with log-linear intensity function (0.2) is

$$l(\beta; x) = \beta \cdot \sum_{x \in x} z(x) - \int_{W} \exp (\beta \cdot z(x)) \, dx. \quad (0.5)$$

Here the integral can be approximated by the Berman and Turner (1992) device, and the MLE can be obtained by spatstat. Asymptotic properties of the MLE are studied in Rathbun and Cressie (1994) and Waagepetersen (2007b).

In general, apart from various parametric models of mixed Poisson processes, the likelihood for Cox process models is intractable. For instance, for the log-Gaussian Cox process driven by (0.3), the covariance function $c(x)$ and hence the distribution of the underlying stationary Gaussian process $\Phi$ may depend on some parameter $\gamma$, and the likelihood function is given by

$$L(\theta; x) = E_{\theta} \left[ \exp \left( -\int_{W} \Lambda(x) \, dx \right) \prod_{x \in x} \Lambda(x) \right]$$
where the expectation is with respect to the distribution of $\Lambda$ with parameter $\theta = (\beta, \gamma)$. This depends on the Gaussian process in such a complicated way that $L(\theta; \mathbf{x})$ is not expressible on closed form. Similarly, the likelihood function for the unknown parameter $\theta = (\beta, \omega, \sigma)$ of the shot-noise Cox process driven by (0.4) is not expressible on closed form.

Møller and Waagepetersen (2003, 2007) discuss various ways of performing maximum likelihood inference based on a missing data MCMC approach (see also Geyer (1999) though this reference is not specifically about Cox processes). Since these methods are rather technical, Section 0.1.3.7 discusses simpler ways of making inference for Cox processes, using moment results as given below.

0.1.3.5 Moments A spatial point process $\mathbf{X}$ has $n$th order product density $\rho^{(n)}(x_1, \ldots, x_n)$ if for all non-negative Borel functions $q(x_1, \ldots, x_n)$ defined for $x_1, \ldots, x_n \in \mathbb{R}^2$,

$$E \sum_{x_1, \ldots, x_n \in \mathbf{X}} q(x_1, \ldots, x_n) = \int \cdots \int q(x_1, \ldots, x_n) \rho^{(n)}(x_1, \ldots, x_n) \, dx_1 \cdots dx_n$$

where the $\neq$ over the summation sign means that the points $x_1, \ldots, x_n$ are pairwise distinct. See e.g. Stoyan et al. (1995). For pairwise distinct $x_1, \ldots, x_n$, we may interpret $\rho^{(n)}(x_1, \ldots, x_n) \, dx_1 \cdots dx_n$ as the probability of observing a point in each of $n$ infinitesimal small regions containing $x_1, \ldots, x_n$ and of areas $dx_1, \ldots, dx_n$, respectively. Of most interest are the intensity function $\rho(x) = \rho^{(1)}(x)$ and the pair correlation function $g(x, y) = \rho^{(2)}(x, y)/(\rho(x)\rho(y))$ (provided $\rho(x)\rho(y) > 0$).

For a Poisson process, Sliвиняк-Mecke’s formula (which is later given by (0.13)) implies $\rho^{(n)}(x_1, \ldots, x_n) = \rho(x_1) \cdots \rho(x_n)$, and so $g(x, y) = 1$. For a general point process $\mathbf{X}$, at least for small distances $\|x - y\|$, we may expect $g(x, y) > 1$ in the case where realizations of $\mathbf{X}$ as compared to a Poisson process tend to be aggregated point patterns (e.g. as in a cluster cluster), and $g(x, y) < 1$ in the case where realizations tend to be regular point patterns (e.g. caused by a repulsion between the points).

For a Cox process driven by $\Lambda$,

$$\rho^{(n)}(x_1, \ldots, x_n) = E[\Lambda(x_1) \cdots \Lambda(x_n)]. \quad (0.6)$$

This takes a nice form for a log-Gaussian Cox process (Møller et al., 1998) where in particular $\rho(x) = \exp(\beta \cdot \mathbf{z}(x))$ (since the Gaussian process $\Phi$ is assumed to be stationary with mean $-c(0)/2$ and $g(x, y) = \exp(c(x - y))$). This shows that the distribution of $(\mathbf{X}, \Phi)$ is specified by $(\rho, g)$, where $g(x, y) = g(x - y)$ (with a slight abuse of notation) is translation invariant and as one may expect, $g(x) > 1$ if and only if $c(x) > 0$. For the shot-noise Cox process given by (0.4), we obtain from (0.6) and Campbell’s theorem (which is later given by (0.13)) that also $\rho(x) = \exp(\beta \cdot \mathbf{z}(x))$. For the modified Thomas process, we obtain from (0.6)
and Slivnyak-Mecke’s formula that \( g(x, y) = g(r) \) depends only on the distance \( r = \|x - y\| \)
where
\[
g(r) = 1 + \exp \left( -r^2 / (4\sigma^2) \right) / (4\pi \omega \sigma^2)
\]
(0.7)
showing that \( g(r) > 1 \) is a decreasing function.

0.1.3.6 Summary statistics and residuals

Exploratory analysis and validation of fitted spatial point process models are usually based on various summary statistics and residuals. In this section we focus mostly on summary statistics and residuals associated to the first and second order moment properties. These apply for many point process models which are not necessarily stationary, including the inhomogeneous Poisson and Cox process models discussed in this contribution, and the Gibbs point process models considered later in Section 0.1.4.

First order properties: In the case of a spatial point process with constant intensity function on \( W \), we naturally use the estimate \( \hat{\rho} = \frac{n(x)}{|W|} \). This is in fact the maximum likelihood estimate if \( X \) is a homogeneous Poisson process on \( W \). In the inhomogeneous case, \( \hat{\rho}(x) = 1 + \frac{c(x)}{\sum_{y \in W} k_0((x - y)/b)} \), \( x \in W \)

(0.8)
is a non-parametric kernel estimate (Diggle, 1985), where \( k_0 \) is a density function with respect to Lebesgue measure, \( b > 0 \) is a user-specified parameter, and \( c(x) = \int_W k_0((x - y)/b) dy \) is an edge correction factor ensuring that \( \int_W \hat{\rho}(x) dx \) is an unbiased estimate of \( \int_W \rho(x) dx \). The kernel estimate is usually sensitive to the choice of the bandwidth \( b \), while the choice of kernel \( k_0 \) is less important.

Various kinds of residuals may be defined (Baddeley et al., 2005; Waagepetersen, 2005). Due to space limitations let us just mention the smoothed raw residual field obtained as follows. By Campbell’s theorem, if we replace \( x \) in (0.8) by \( X \cap W \), then \( \hat{\rho}(x) \) has mean \( \int_W \rho(y)k_0((x - y)/b) dy / c(x) \). Suppose we have fitted a parametric model with parameter \( \theta \) and estimate \( \hat{\theta} \), where the intensity function \( \rho_\theta \) may depend on \( \theta \). Then the smoothed raw residual field is given by

\[
s(x) = \hat{\rho}(x) - \frac{1}{c(x)} \int_W \rho_\theta(y)k_0((x - y)/b) dy, \quad x \in W.
\]
(0.9)
Positive/negative values of \( s \) suggest that the fitted model under/overestimates the intensity function. In spatstat the residual field is represented as a greyscale image and a contour plot. See also Figure 0.5 in Section 0.1.3.8.

Second order properties: Assume that the pair correlation function is invariant under translations, i.e. \( g(x, y) = g(x - y) \). This assumption is called second order intensity reweighted stationarity (Baddeley et al., 2000) and it is satisfied for the Poisson process, the log-Gaussian Cox, and shot-noise Cox processes studied in this contribution, cf. Section 0.1.3.5.
The inhomogeneous $K$ and $L$-functions (Baddeley et al., 2000) are defined by

$$K(r) = \int_{\|x\| \leq r} g(x) \, dx, \quad L(r) = \sqrt{K(r)/\pi}, \quad r > 0.$$ 

In the stationary case we obtain Ripley’s $K$-function (Ripley, 1976; Ripley, 1977) with the interpretation that $\rho K(r)$ is the expected number of points within distance $r$ from the origin $o$, where the expectation is with respect to the so-called reduced Palm distribution at $o$ (intuitively, this is the conditional distribution given that $X$ has a point at $o$). For a Poisson process, $L(r) = r$ is the identity, which is one reason why one often uses the $L$-function (see also Besag (1977b)).

For a log-Gaussian Cox process, we may obtain $K(r)$ by numerical methods. For the modified Thomas process, (0.7) implies that

$$K(r) = \pi r^2 + \left(1 - \exp\left(-r^2/(4\sigma^2)\right)\right)/\omega. \quad (0.10)$$

In order to investigate for a possible anisotropy, directional $K$-functions have been constructed (Stoyan and Stoyan, 1995; Brix and Möller, 2001), while Mugglestone and Renshaw (1996) use a method as part of two-dimensional spectral analysis. Frequently, $g$ is assumed to be invariant under rotations, i.e. $g(x, y) = g(\|x - y\|)$, in which case $K$ and $g$ are in a one-to-one correspondence, and it is usually easy to estimate and investigate a plot of $g$. Non-parametric kernel estimation of $g$ is discussed in Stoyan and Stoyan (2000); it is sensitive to the choice of bandwidth and a truncation near zero is needed.

An unbiased estimate of $K(r)$ is given by

$$\hat{K}(r) = \sum_{x \neq y \in X} \frac{\mathbb{I}[\|x - y\| \leq r]}{\rho(x)\rho(y)|W \cap (W + x - y)|} \quad (0.11)$$

where $W + x$ denotes $W$ translated by $x$, and $|W \cap (W + x - y)|$ is an edge correction factor, which is needed since we sum over all pairs of points observed within $W$. In practice we need to plug in an estimate of $\rho(x)\rho(y)$, and denote the resulting estimate of $K$ by $\hat{K}$. It may be given by a parametric estimate $\hat{\rho}(x)\hat{\rho}(y)$ or by a non-parametric estimate, e.g. $\hat{\rho}(x)\hat{\rho}(y)$ from (0.8) (but see also Stoyan and Stoyan (2000) in the stationary case and Baddeley et al. (2000) in the inhomogeneous case). Using the parametric estimate seems more reliable, since the degree of clustering exhibited by $\hat{K}$ may be very sensitive to the choice of bandwidth if (0.8) is used.

The non-parametric estimate $\hat{K}$ for the tropical rain forest trees (Figures 0.1-0.2) obtained with a parametric estimate of the intensity function is shown in Figure 0.3 (Section 0.1.3.7 describes the estimation procedure). The plot also shows theoretical $K$-functions for fitted log Gaussian Cox, modified Thomas, and Poisson processes, where all three processes share the same intensity function (see again Section 0.1.3.7). The trees seem to form a clustered point pattern, since $\hat{K}$ is markedly larger than the theoretical $K$-function for a Poisson process.
Fig. 0.3. Estimated $K$-function for tropical rain forest trees and theoretical $K$-functions for fitted Thomas, log Gaussian Cox, and Poisson processes.

Further summaries and confidence bounds: Other summary statistics than those based on first and second order moment properties usually assume stationarity of the point process $X$. This includes the $F, G, J$-functions below, which are based on interpoint distances. Also summary statistics based on higher-order moment properties exist (Stoyan and Stoyan, 1995; Möller et al., 1998; Schladitz and Baddeley, 2000).

Suppose that $X$ is stationary with intensity $\rho > 0$. The empty space function $F$ and the nearest-neighbour function $G$ are the cumulative distribution functions of the distance from respectively an arbitrary location to the nearest point in $X$ and a ‘typical’ point in $X$ (i.e. with respect to the reduced Palm distribution at $o$) to its nearest-neighbour point in $X$. For a stationary Poisson process, $F(r) = G(r) = 1 - \exp(-\pi r^2)$. In general, at least for small distances, $F(r) < G(r)$ indicates aggregation and $F(r) > G(r)$ indicates regularity. Van Lieshout and Baddeley (1996) study the nice properties of the $J$-function defined by $J(r) = (1 - G(r))/(1 - F(r))$ for $F(r) < 1$. Non-parametric estimation of $F$ and $G$ accounting for edge effects is straightforward, see e.g. Stoyan et al. (1995), and combining the estimates we obtain an estimate of $J$.

Finally, estimates $\hat{g}(r), \hat{K}(r), \hat{F}(r), \ldots$ may be supplied with confidence intervals for each value of $r$ obtained by simulations $\hat{g}_i(r), \hat{K}_i(r), \hat{F}_i(r), \ldots$, $i = 1, 2, \ldots$ under an estimated model, see e.g. Möller and Waagepetersen (2003).
Composite likelihood and minimum contrast estimation. The intensity functions of the log-linear Poisson process, the log-Gaussian, and shot-noise Cox processes in Section 0.1.3.2 are of the same form, viz. \( \rho_\beta(x) = \exp(\beta \cdot \mathbf{z}(x)) \), cf. Section 0.1.3.5. Their pair correlation functions are different, where \( g_\psi \) depends on a parameter \( \psi \) which is usually assumed to be variation independent of \( \beta \in \mathbb{R}^{p+1} \), e.g. \( \psi = (\omega, \sigma) \in (0, \infty)^2 \) in the case of the modified Thomas process, cf. (0.7). Following Möller and Waagepetersen (2007) we may estimate \( \beta \) by maximizing a composite likelihood function based on \( g_\psi \) or its corresponding \( K \)-function \( K_\psi \), given by e.g. (0.10).

The composite likelihood function is derived as a certain limit of a product of marginal likelihoods. Let \( C_i \), \( i \in I \), be a finite partitioning of \( W \) into disjoint cells \( C_i \) of small areas \( |C_i| \), and define \( N_i = 1[\mathbf{X} \cap C_i \neq \emptyset] \) and \( p_i(\theta) = P_\theta(N_i = 1) \approx \rho_\theta(u_i)|C_i| \), where \( u_i \) denotes a representative point in \( C_i \). Under mild conditions the limit of \( \log \prod_i p_i(\theta)^{N_i}(1-p_i(\theta))^{(1-N_i)} \) becomes (0.5); this is the log composite likelihood function.

For the rain forest trees (Figures 0.1-0.2) we obtain the maximum composite likelihood estimate \((\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = (-4.989, 0.021, 5.842)\) (under the Poisson model this is the MLE). Assuming asymptotic normality (Waagepetersen, 2007b) 95% confidence intervals for \( \beta_1 \) and \( \beta_2 \) under the fitted shot-noise Cox process are \([-0.018, 0.061]\) and \([0.885, 10.797]\), respectively, while much more narrow intervals are obtained under the fitted Poisson process \([0.017, 0.026]\) and \([5.340, 6.342]\). For instance, \( \psi \) may be estimated by minimizing the contrast

\[
\int_a^b \left( K(r)^\alpha - K_\psi(r)^\alpha \right)^2 \, dr
\]

where typically \( 0 \leq a < b < \infty \) and \( \alpha > 0 \) are chosen on an ad hoc basis, see e.g. Diggle (2003) and Møller and Waagepetersen (2003). For the rain forest trees, replacing \( \rho(x)\rho(y) \) in (0.11) by the parametric estimate of \( \rho(x)\rho(y) \) obtained from above and taking \( a = 0, b = 100, \) and \( \alpha = 1/4 \), we obtain \((\hat{\omega}, \hat{\sigma}) = (8 \times 10^{-5}, 20)\). The estimated theoretical \( K \)-functions are shown in Figure 0.3.

These ‘simulation-free’ estimation procedures are fast and computationally easy, but the disadvantage is that we have to specify tuning parameters such as \( a, b, \alpha \). Theoretical properties of maximum composite likelihood estimators are investigated in Waagepetersen (2007b) and Waagepetersen and Guan (2007), and of the minimum contrast estimators in Heinrich (1992).

Simulation-based Bayesian inference. A Bayesian approach often provides a flexible framework for incorporating prior information and analyzing Poisson and Cox process models.

As an example, consider Figure 0.4 which shows the locations of just one species (called seeder 4 in Illian et al. (2007)) from a much larger data set where
the locations of 6378 plants from 67 species on a 22 m by 22 m observation window $W$ in the south western area of Western Australia have been recorded (Armstrong, 1991). The plants have adapted to regular natural fires, where resprouting species survive the fire, while seeding species die in the fire but the fire triggers the shedding of seeds, which have been stored since the previous fire. As argued in more detail in Illian et al. (2007) it is therefore natural to model the locations of the reseeding plants conditionally on the locations of the resprouting plants.

![Fig. 0.4. Locations of Leucopogon conostephioides plants observed within a 22 × 22 m window.](image)

In the sequel we consider the 19 most dominant (influential) species of resprouters, with $y_1, \ldots, y_{19}$ the observed point patterns of resprouters as given in Figure 1 in Illian et al. (2007), and define covariates by

$$z_i(x) = \sum_{y \in y_i: \|x-y\| \leq \kappa_{y,i}} \left(1 - \left(\frac{\|x-y\|}{\kappa_{y,i}}\right)^2\right)^2, \quad i = 1, \ldots, 19$$

where $\kappa_{y,i} \geq 0$ is the radius of interaction of the $i$th resprouter at location $y$. Note that we suppress in the notation that $z_i$ depends on all the $\kappa_{y,i}$ with $y \in y_i$.

As usual $X \cap W = x$ denotes the data (here the point pattern of seeder 4), where we assume that $X \cap W$ is a Poisson process with log-linear intensity function (0.2) and parameter $\beta = (\beta_0, \ldots, \beta_{19})$, where $\beta_0$ is an intercept and $\beta_i$ for $i > 0$ controls the influence of the $i$th resprouter. Thus the log-likelihood is of the form (0.5) but with unknown parameter $\theta = (\beta, \kappa)$, where $\kappa$ denotes the collection of all the radii $\kappa_{y,i}$. 
Using a fully Bayesian set up, we treat \( \theta = (\beta, \kappa) \) as a random variable, where Table 1 in Illian et al. (2007) provides some prior information on \( \kappa \). Specifically, following Illian et al. (2007), we assume a priori that the \( \kappa_{y,i} \) and the \( \beta_i \) are mutually independent, each \( \kappa_{y,i} \) follows the restriction of a normal distribution \( N(\mu_i, \sigma_i^2) \) to \([0, \infty)\), where \((\mu_i, \sigma_i^2)\) is chosen so that under the unrestricted normal distribution the range of the zone of influence is a central 95\% interval, and each \( \beta_i \) is \( N(0, \sigma^2) \)-distributed, where \( \sigma^2 = 8 \). Combining these prior assumptions with the log-likelihood we obtain the posterior density

\[
\pi(\beta, \kappa | x) \propto \exp \left( \frac{\beta_0}{(2\sigma^2)} - \sum_{i=1}^{19} \left\{ \frac{\beta_i^2}{(2\sigma^2)} + \sum_{y \in y_i} (\kappa_{y,i} - \mu_i)^2/(2\sigma_i^2) \right\} \right) \\
\times \exp \left( \beta \cdot \sum_{x \in X} z(x) - \int_W \exp(\beta \cdot z(u) \, du) \right), \quad \beta_i \in \mathbb{R}, \kappa_{y,i} \geq 0 \quad (0.12)
\]

(suppressing in the notation that \( z \) depends on \( \kappa \) and we have conditioned on \( y_1, \ldots, y_{19} \) in the posterior distribution).

Monte Carlo estimates of various marginal posterior distributions are calculated using simulations from (0.12) obtained by a hybrid MCMC algorithm (see e.g. Robert and Casella (1999)) where we alter between updating \( \beta \) and \( \kappa \) using random walk Metropolis updates (for details, see Illian et al. (2007)). For example, a large (small) value of \( P(\beta_i > 0 | x) \) indicates a positive/attractive (negative/repulsive) association to the \( i \)th resprouter, see Figure 2 in Illian et al. (2007). The model can be checked following the idea of posterior predictive model assessment (Gelman et al., 1996), comparing various summary statistics with their posterior predictive distributions. The posterior predictive distribution of statistics depending on \( X \) (and possibly also on \( (\beta, \kappa) \)) is obtained from simulations: we generate a posterior sample \((\beta^{(j)}, \kappa^{(j)})\), \( j = 1, \ldots, m \), and for each \( j \) ‘new data’ \( x^{(j)} \) from the conditional distribution of \( X \) given \((\beta^{(j)}, \kappa^{(j)})\). For instance, the grey scale plot in Figure 0.5 is a ‘residual’ plot based on quadrant counts. We divide the observation window into 100 equally sized quadrants and count the number of seeder 4 plants within each quadrant. The grey scales reflect the probabilities that counts drawn from the posterior predictive distribution are less or equal to the observed quadrant counts where dark means high probability. The stars mark quadrants where the observed counts are ‘extreme’ in the sense of being either below the 2.5\% quantile or above the 97.5\% quantile of the posterior predictive distribution. Figure 0.5 does not provide evidence against our model. A plot based on the \( L \)-function and the posterior predictive distribution is also given in Illian et al. (2007); again there is no evidence against our model.

Møller and Waagepetersen (2007) discuss another Bayesian analysis for the rain forest trees, using a shot-noise Cox process model. They consider the Poisson process \( Y \) of mother points from (0.4) as one part of the prior (more precisely, to obtain a finite process, \( Y \) is restricted to a bounded region \( B \supseteq W \) as discussed in Section 0.1.3.3); remaining unknown parameters are denoted \( \theta \), and a certain prior is imposed on \( \theta \). Simulations from the posterior are again obtained.
by a hybrid MCMC algorithm, where one type of update involves conditional simulation of $Y$ given the data and $\theta$, using the Metropolis-Hastings birth-death algorithm in Section 0.1.4.7 (see also Møller (2003)).

For simulation-based Bayesian inference for a log-Gaussian Cox process, we need among other things to make conditional simulation of the Gaussian process $\Phi$ in (0.3) given the data and $\theta$. This may be done by a Langevin-Hastings algorithm (Besag, 1994; Roberts and Tweedie, 1996) as detailed in Møller et al. (1998) and Møller and Waagepetersen (2003). Promising results in Rue and Martino (2005) suggest that it may be possible to compute accurate approximations of posterior distributions without MCMC.

0.1.4 Inference for Gibbs point processes

While Cox processes provide flexible models for aggregation or clustering in a point pattern, Gibbs point processes provide flexible models for regularity or repulsion (Van Lieshout, 2000; Møller and Waagepetersen, 2003). Though the density has an unknown normalizing constant (Section 0.1.4.1), likelihood inference based on MCMC methods is easier for parametric Gibbs point process models (Sections 0.1.4.2 and 0.1.4.4) than for Cox process models. On the other
hand, Bayesian inference is more complicated, involving an auxiliary variable MCMC method and perfect simulation based on a technique called dominated coupling from the past (Sections 0.1.4.6 and 0.1.4.7).

An example of a regular point pattern is shown in Figure 0.6. The data set is actually a marked point pattern, with points given by the tree locations and marks by the stem diameters, where the discs may be thought of as ‘influence zones’ of the trees. Figure 0.7 shows estimates of \( F, G, J \) and provides evidence of regularity in the point pattern, which in fact to a large extent is due to forest management but may also be caused by repulsion between the trees. The data set has been further analyzed using point process methods by Fiksel (1984), Penttinen et al. (1992), Goulard et al. (1996), and Möller and Waagepetersen (2003, 2007).

![Figure 0.6](image)

**Fig. 0.6.** Norwegian spruces observed in a 56 × 38 m window in Tharandter Forest, Germany. The radii of the 134 discs equal 5 times the stem diameters.

0.1.4.1 Gibbs point processes Gibbs point processes arose in statistical physics as models for interacting particle systems (see e.g. Ruelle (1969)), and they have important applications in stochastic geometry and spatial statistics (see e.g. Baddeley and Möller (1989), Van Lieshout (2000), and Stoyan et al. (1995)). They possess certain Markov properties which are useful for ‘local computations’ and to account for edge effects. Summary statistics for Gibbs point processes, including the intensity function \( \rho \), are in general not expressible on closed form. Instead the Papangelou conditional intensity (Papangelou, 1974) becomes the appropriate tool.

**Definition of conditional intensity** Below we define the Papangelou conditional intensity \( \lambda(x; x) \) for a general point process \( X \), assuming the process exists; the question of existence of Gibbs point processes is discussed later. Here \( x \subset \mathbb{R}^2 \) denotes a locally finite point configuration and \( x \in \mathbb{R}^2 \).

By the Georgii-Nguyen-Zessin formula (Georgii, 1976; Nguyen and Zessin,
Fig. 0.7. Left to right: estimated $F, G, J$-functions for the Norwegian spruces (solid lines) and 95% envelopes calculated from simulations of a homogeneous Poisson process (dashed lines) with expected number of points equal to the observed number of points. The long-dashed curves show the theoretical values of $F, G, J$ for a Poisson process.

1979), $X$ has Papangelou conditional intensity $\lambda$ if this is a non-negative Borel function such that

$$E \sum_{x \in X} h(x, X \setminus \{x\}) = \int E[\lambda(x; X)h(x, X)] \, dx \quad (0.13)$$

for non-negative Borel functions $h$. The GNZ-formula (0.13) is a very general and useful but indirect definition/characterization of the conditional intensity. If $P^X_x$ is the reduced Palm distribution at the point $x$ (intuitively, this is the conditional distribution given that $X$ has a point at $x$) and $P^X_x$ is absolutely continuous with respect to the distribution $P$ of $X$, we may take $\lambda(x; x) = \rho(x) (dP^X_x / P)(x)$ (Georgii, 1976). Thus, for $x \notin x$, we may interpret $\lambda(x; x) \, dx$ as the conditional probability that there is a point of the process in an infinitesimally small region containing $x$ and of area $dx$ given that the rest of the point process coincides with $x$. For a Poisson process, $\lambda(x; x) = \rho(x)$ and (0.13) becomes the Slivnyak-Mecke formula (Mecke, 1967). When $h(x, X \setminus \{x\}) = h(x)$ only depends on its first argument, (0.13) reduces to Campbell’s theorem (e.g. Stoyan et al. (1995)) with $\rho(x) = E[\Lambda(x; X)]$.

Existence and uniqueness of $\lambda$ is often not a trivial issue (Georgii, 1976; Preston, 1976; Georgii, 1988). In fact two infinite point processes may share the same Papangelou conditional intensity (even in the stationary case); this phenomenon is known in statistical physics as phase transition. By (0.13), for a Cox process driven by $\Lambda$, we can take $\lambda(x; X) = E[\Lambda(x; X)|X]$, but this conditional expectation is in general not expressible on closed form. As explained later, the GNZ-formula is more useful in connection to Gibbs point processes.

In particular, the conditional intensity exists for a finite point process $X$ defined on a bounded region $B \in \mathcal{B}$, assuming this has an density $f$ with respect to the unit rate Poisson process on $B$, and $f$ is hereditary, i.e.

$$f(y) > 0 \text{ whenever } f(x) > 0, \ y \subset x \subset B, \ x \text{ finite.} \quad (0.14)$$
Inference 17

Then we can take
\[ \lambda(x; x) = \frac{f(x \cup \{x\})}{f(x \setminus \{x\})} \quad (0.15) \]
if \( f(x \setminus \{x\}) > 0 \), and \( \lambda(x; x) = 0 \) otherwise. The precise definition of \( \lambda(x; x) \) when \( x \in \mathbf{x} \) is not that important, and (0.15) just covers this case for completeness. In fact \( \lambda(x; x) \) is unique up to a Lebesgue null set of points \( x \) times a Poisson null set of point configurations \( \mathbf{x} \). Note that \( \lambda(x; x) = \lambda(x; x\setminus\{x\}) \), and (0.14) implies that \( f \) and \( \lambda \) are in a one-to-one correspondence. Often we specify \( f(x) \propto h(x) \) only up to proportionality, and call \( h \) an unnormalized density. Here we assume that \( h \) is hereditary and integrable with respect to the unit rate Poisson process on \( B \), in which case we can replace \( f \) by \( h \) in (0.15). These conditions are ensured by local stability, which means that for some finite constant \( K \),

\[ h(x \cup \{x\}) \leq K h(x) \quad \text{for all finite } x \subset B \text{ and } x \in B \setminus x. \quad (0.16) \]

Then \( \lambda(x; x) \leq K \) is uniformly bounded. A weaker condition ensuring integrability of \( h \) is Ruelle stability (Ruelle, 1969). We restrict attention to local stability, since this property is usually satisfied for Gibbs point processes used in stochastic geometry, and it implies many desirable properties as discussed in Section 0.1.4.7.

**Definition and Markov properties of Gibbs point processes** Suppose that \( R > 0 \) is a given parameter and \( U(x) \in (-\infty, \infty] \) is defined for all finite \( x \subset \mathbb{R}^2 \) such that \( U(x) = 0 \) if \( x \) contains two points \( R \) units apart. Then a point process \( \mathbf{X} \) is a Gibbs point process (also called a canonical ensemble in statistical physics and a Markov point process in spatial statistics) with potential \( U \) if it has a Papangelou conditional intensity of the form

\[ \lambda(x; \mathbf{x}) = \exp \left( - \sum_{y \in \mathbf{x} \cap B_R(x)} U(y \cup \{x\}) \right) \]

for all locally finite \( \mathbf{x} \subset \mathbb{R}^2 \) and \( x \in \mathbb{R}^2 \setminus \mathbf{x} \), where \( B_R(x) \) is the closed disc of radius \( R \) centered at \( x \). Note that the conditional intensity is zero if \( U(y \cup \{x\}) = \infty \). It satisfies a local Markov property: \( \lambda(x; \mathbf{x}) \) depends on \( \mathbf{x} \) only through \( \mathbf{x} \cap \partial B_R(x) \), i.e. the \( R \)-close neighbours in \( \mathbf{x} \) to \( x \). Usually \( R \) is chosen as small as possible, in which case it is called the range of interaction.

An equivalent characterization of a Gibbs point process is given by a so-called local specification (Preston, 1976). This means that a spatial Markov property should be satisfied: for any bounded \( B \in \mathcal{B} \), the conditional distribution of \( \mathbf{X} \cap B \) given \( \mathbf{X} \setminus B \) agrees with the conditional distribution of \( \mathbf{X} \cap B \) given \( \mathbf{X} \cap \partial B \), where \( \partial B = \{x \in \mathbb{R}^2 \setminus B : B(x) \cap B \neq \emptyset\} \) is the \( R \)-close neighbourhood to \( B \). It also means that the conditional density of \( \mathbf{X} \cap B \) given \( \mathbf{X} \cap \partial B = \partial \mathbf{x} \) is of the form

\[ f_B(\mathbf{x}|\partial \mathbf{x}) = \frac{1}{c_B(\partial \mathbf{x})} \exp \left( - \sum_{y \in \mathbf{x} \setminus \partial \mathbf{x} \neq \emptyset} U(y \cup \partial \mathbf{x}) \right) \quad (0.17) \]
with respect to the unit rate Poisson process on $B$. Here $c_B(\partial x)$ is a normalizing constant (called the partition function in statistical physics), which in general is not expressible on closed form.

Conditions ensuring existence of an infinite Gibbs point process, uniqueness or non-uniqueness (the latter is known as phase transition in statistical physics), and stationarity are rather technical (Georgii, 1976; Preston, 1976; Georgii, 1988). In general closed form expressions for $\rho(n)$ and other summary statistics are unknown (even in the stationary case or in the finite case).

For a finite Gibbs point process $X$ defined on a bounded region $B \in \mathcal{B}$, the density is of the form

$$f(x) = \frac{1}{c} \exp \left( - \sum_{y \subseteq x, y \neq \emptyset} U(y) \right)$$

(0.18)

corresponding to the case of (0.17) with ‘free boundary condition’ $\partial x = \emptyset$. The celebrated Hammersley-Clifford theorem (Ripley and Kelly, 1977) states that for a finite point process on $B$ with hereditary density $f$, the local Markov property is equivalent to (0.18).

The spatial Markov property can be used to account for edge effects when the Gibbs process is observed within a bounded window $W$ but it extends outside $W$. Let $W_{\Theta R} = \{ x \in W : B_R(x) \subseteq W \}$ be the $R$-clipped window. Conditional on $X \cap \partial W_{\Theta R} = \partial x$, we have that $X \cap W_{\Theta R}$ is independent of $X \setminus W$. The conditional density of $X \cap W_{\Theta R}\mid X \cap \partial W_{\Theta R} = \partial x$ is given by (0.17) with $B = W_{\Theta R}$, and combining this with (0.15) we see that the conditional intensity $\lambda_{W_{\Theta R} \cap \partial x}(x; x) \mid X \cap \partial W_{\Theta R} = \partial x$ agrees with $\lambda$, that is

$$\lambda_{W_{\Theta R} \cap \partial x}(x; x) = \lambda(x; x \cup \partial x), \quad x \subset W_{\Theta R}, \ x \in W_{\Theta R} \setminus x.$$ (0.19)

### 0.1.4.2 Modelling conditional intensity

Usually parametric models of a Gibbs point process with a fixed interaction radius $R$ have a conditional intensity of a log-linear form

$$\log \lambda_0(x; x) = \theta \cdot t(x; x) = \sum_{i=1}^{k} \theta_i \cdot t_i(x; x)$$

(0.20)

where $\theta = (\theta_1, \ldots, \theta_k)$ is a real parameter vector (possibly infinite parameter values of some $\theta_i$ are also included as exemplified below for a hard core process), $t(x; x) = (t_1(x; x), \ldots, t_k(x; x))$ is a real vector function, where $t_i(x; x)$ of the same dimension as $\theta_i$, $i = 1, \ldots, k$, and $\cdot$ is the usual inner product.

We say that a Gibbs point process has interactions of order $k$, if $k$ is the smallest integer so that the potential $U(x)$ does not vanish if $n(x) \leq k$. Then we usually assume that

$$U(x) = -\theta_i \cdot V(x) \quad \text{if} \ n(x) = i \leq k$$

(0.21)

where $V(x)$ is a vector function with non-negative components, and
The first order term $V(\{x\})$ is usually constant one or it is given by a vector of covariates $V(\{x\}) = \mathbf{z}(x)$, where $\theta_1 = \beta$ consists of regression parameters (Ogata and Tanemura, 1989), while the higher order terms $V(x) (n(x) > 0)$ are usually one-dimensional.

We often consider pairwise interaction processes with repulsion between the points, i.e. $\theta_2 \leq 0$ and $\theta_n = 0$ whenever $n \geq 3$, where in many cases $V(\{x, y\}) = V(\|x - y\|)S$ is a real decreasing function of $\|x - y\|$. A special case is the Strauss process (Strauss, 1975; Kelly and Ripley, 1976) where $V(\{x\}) = 1, V(\{x, y\}) = \mathbb{1}[\|x - y\| \leq R]$, and $(\theta_1, \theta_2) \in \mathbb{R} \times [-\infty, 0]$; if $\theta_2 = -\infty$, we use in (0.21) the convention that $-\infty \times 0 = 0$ and $-\infty \times 1 = -\infty$, and call $R$ a hard core parameter since all points in the process are at least $R$-units apart from each other. Figure 0.8 shows perfect simulations (for details, see Section 0.1.4.7) of different Strauss processes restricted to the unit square.

![Simulation of a Strauss process on the unit square](image)

**Fig. 0.8.** Simulation of a Strauss process on the unit square (i.e. $V(\{x\}) = 1$ if $x \in [0,1]^2$ and $V(\{x\}) = 0$ otherwise) when $R = 0.05, \theta_1 = \log(100)$, and $\theta_2 = 0, -\log 2, -\infty$ (from left to right).

The Widom-Rowlinson model (Widom and Rowlinson, 1970) (or area-interaction process, Baddeley and Van Lieshout (1995)) is given by

$$
\log \lambda_\theta(x; \mathbf{x}) = \beta - \psi |B_x(R) \setminus \bigcup_{y \in \mathbf{x}} B_y(R)|
$$

where $\psi, \beta \in \mathbb{R}$. This is clearly of the form (0.20), and for a finite area-interaction process the corresponding density is proportional to

$$
\exp \left( \beta n(\mathbf{x}) - \psi \left| \bigcup_{y \in \mathbf{x}} B_y(R) \right| \right).
$$

It follows that this is a Markov point process with interaction range $R$, but the Hammersley-Clifford representation (0.18) is not so useful: using the inclusion-exclusion formula on $| \bigcup_{y \in \mathbf{x}} B_y(R) |$, we obtain this representation, but the expression for the potential is intricate. For $\psi \neq 0$, the process has interactions of all orders, since the $t_i(x)$-terms in (0.22) do not vanish for any $i \geq 1$ and the
corresponding \( \theta_i = \psi \) is non-zero for all \( i \geq 2 \). The model is said to be attractive if \( \psi > 0 \) and repulsive if \( \psi < 0 \), since \( \lambda_\theta(x; \mathbf{x}) \) is an increasing respective decreasing function of \( \mathbf{x} \). Phase transition happens for all sufficient numerical large values of \( \psi > 0 \) (Ruelle, 1971; Häggström et al., 1999).

Further specific examples of Gibbs point process models are given in Van Lieshout (2000) and the references therein.

Consider again Figure 0.6. The conditional intensity for a Norwegian spruce with a certain influence zone should depend not only on the positions but also on the influence zones of the neighbouring trees. A tree with influence zone given by the disc \( B_r(x) \), where \( x \) is the spatial location of the tree and \( r \) is the influence zone radius, is treated as a point \( (x, r) \) in \( \mathbb{R}^2 \times (0, \infty) \). We assume that the influence zone radii belong to a bounded interval \( M = [a, b] \), where \( a \) and \( b \) are estimated by the minimal and maximal observed influence zone radii, and we divide \( M \) into six disjoint subintervals of equal size. Confining ourselves to a pairwise interaction process with repulsion, we let

\[
\lambda_\theta((x, r); \mathbf{x}) = \exp \left( \beta(r) + \psi \sum_{(x', r') \in \mathbf{x}} |B_r(x) \cap B_{r'}(x')| \right)
\]

where \( \beta(r) = \beta_k \) if \( r \) falls in the \( k \)th subinterval, and \( \theta = (\beta_1, \ldots, \beta_6, \psi) \in \mathbb{R}^6 \times (-\infty, 0] \). This enables modelling the varying numbers of trees in the six different size classes, and the strength of the repulsion between two trees is given by \( \psi \) times the area of overlap between the influence zones of the two trees. However, the interpretation of the conditional intensity is not straightforward — it is for instance not in general a monotone function of \( r \). On the other hand, for a fixed \( (x, r) \), the conditional intensity will always decrease if the neighbouring influence zones increase.

0.1.4.3 Residuals and summary statistics Exploratory analysis and validation of fitted Gibbs point process models can be done using residuals and summary statistics, cf. Figure 0.7. As mentioned theoretical closed form expressions for summary statistics such as \( \rho, g, K, L, F, G, J \) are in general unknown for Gibbs point process models, so one has entirely to rely on simulations. Residuals are more analytical tractable due to the GNZ-formula (0.13), see Baddeley et al. (2005, 2007). These papers discuss in detail how various kinds of residuals can be defined, where \texttt{spatstat} can be used for the log-linear model (0.20). For example, the smoothed raw residual field is now given by

\[
s(x) = \hat{\rho}(x) - \frac{1}{c(x)} \int_W \lambda_\theta(y; \mathbf{x}) k_0(\|x - y\|/b) \, dy,
\]

with a similar interpretation of positive/negative values as in (0.9) but referring to the conditional intensity instead.

0.1.4.4 Simulation-based maximum likelihood inference Ogata and Tanemura (1984) and Penttinen (1984) are pioneering works on simulation-based maximum
likelihood inference for Markov point processes; see also Geyer and Møller (1994) and Geyer (1999). This section considers a parametric Gibbs point process model with finite interaction radius $R$, potential of the form (0.21), and finite interaction of order $k$ (though other cases such as the area-interaction process are covered as well, cf. Section 0.1.4.2). We assume to begin with that $R$ is known.

First, suppose that the process has no points outside the observation window $W$. By (0.18) and (0.21) the log-likelihood
\[ l(\theta; x) = \theta \cdot t(x) - \log c_\theta, \quad t(x) = (t_1(x), \ldots, t_k(x)) \tag{0.24} \]
is of exponential family form (Barndorff-Nielsen, 1978), where
\[ t_i(x) = \sum_{y \subseteq x : n(y) = i} V(y), \quad i = 1, \ldots, k. \]
Thus the score function (the first derivative of $l$ with respect to $\theta$) and observed information (the second derivative) are
\[ u(\theta) = t(x) - E_\theta t(X), \quad j(\theta) = \text{Var}_\theta t(X), \]
where $E_\theta$ and $\text{Var}_\theta$ denote expectation and variance with respect to $X$ with parameter $\theta$. Consider a fixed reference parameter value $\theta_0$. The score function and observed information may then be evaluated using the importance sampling formula
\[ E_\theta q(X) = E_{\theta_0} \left[ q(X) \exp \left( (\theta - \theta_0) \cdot t(X) \right) \right] / (c_\theta / c_{\theta_0}) \]
with $q(X)$ given by $t(X)$ or $t(X)^2 t(X)$. The importance sampling formula also yields
\[ c_\theta / c_{\theta_0} = E_{\theta_0} \left[ \exp \left( (\theta - \theta_0) \cdot t(X) \right) \right]. \]
Approximations of the log likelihood ratio $l(\theta; x) - l(\theta_0; x)$, score, and observed information are then obtained by Monte Carlo approximation of the expectations $E_{\theta_0}[\cdot]$ using MCMC samples from the process with parameter $\theta_0$, see Section 0.1.4.7. The path sampling identity (Gelman and Meng, 1998) provides an alternative and often numerically more stable way of computing a ratio of normalizing constants.

Second, if $X$ may have points outside $W$, we have to correct for edge effects. By (0.17) the log likelihood function based on the conditional distribution $X \cap W_{\cap R} X \cap \partial W_{\cap R} = \partial x$ is of similar form as (0.24) but with $c_\theta$ depending on $\partial x$. Thus likelihood ratios, score, and observed information may be computed by analogy with the case above. Alternatively, if $X$ is contained in a bounded region $S \supset W$, the likelihood function based on the distribution of $X$ may be computed using a missing data approach. This is a more efficient and complicated approach, see Geyer (1999) and Møller and Waagepetersen (2003). Yet other approaches for handling edge effects are discussed in Møller and Waagepetersen (2003).

For a fixed $R$, the approximate (conditional) likelihood function can be maximized with respect to $\theta$ using Newton-Raphson updates. Frequently the Newton-Raphson updates converge quickly, and the computing times for obtaining a MLE
are modest. MLE’s of $R$ are often found using a profile likelihood approach, since
the likelihood function is typically not differentiable and log concave as a function
of $R$.

Asymptotic results for MLE’s of Gibbs point process models are established
in Mase (1991) and Jensen (1993) but under restrictive assumptions of stationar-
ity and weak interaction. According to standard asymptotic results, the inverse
observed information provides an approximate covariance matrix of the MLE,
but if one is suspicious about the validity of this approach, an alternative is to
use a parametric bootstrap, see e.g. Møller and Waagepetersen (2003).

For the overlap interaction model (0.23), Møller and Waagepetersen (2003)
computed MLE’s using both missing data and conditional likelihood approaches,
where the conditional likelihood approach is based on the trees with locations
in $W \ominus 2b$, since trees with locations outside $W$ do not interact with trees located
inside $W \ominus 2b$. The conditional MLE is given by

$$(\hat{\beta}_1, \ldots, \hat{\beta}_6) = (-1.02, -0.41, 0.60, -0.67, -0.58, -0.22), \quad \hat{\psi} = -1.13.$$ 

Confidence intervals for $\psi$ obtained from the observed information and a para-
metric bootstrap are $[-1.61, -0.65]$ and $[-1.74, -0.79]$, respectively. The fitted
overlap interaction process seems to capture well the second order characteristics
for the point pattern of tree locations, see Figure 0.9.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig09}
\caption{Model assessment for Norwegian spruces: estimated $L(r) - r$ func-
tion for spruces (solid line) and average and 95\% envelopes computed from
simulations of fitted overlap interaction model (dashed lines).}
\end{figure}

0.1.4.5 Pseudo-likelihood The maximum pseudo likelihood estimate (MPLE)
is a simple and computational fast but less efficient alternative to the MLE.

First, consider a finite point process with no points outside the observa-
tion window $W$. Let $C_i, i \in I$, be a finite partitioning of $W$ into disjoint cells
C_i of small areas |C_i|, and define N_i = I[X ∩ C_i ≠ ∅] and p_i(θ) = P_θ(N_i = I[X \ C_i]) ≈ λ_θ(u_i, X \ C_i)|C_i|, where u_i denotes a representative point in C_i. Under mild conditions (Besag et al., 1982; Jensen and Møller, 1991) the limit of log \( \prod_i (p_i(θ)/|C_i|)^{N_i}(1 - p_i(θ))^{(1 - N_i)} \) becomes

\[
\sum_{x \in X} \lambda_θ(x, x) - \int_W \lambda_θ(x, x) \, dx
\]  

which is known as the log pseudo-likelihood function (Besag, 1977a). By the GNZ formula (0.13), the pseudo score

\[
s(θ) = \sum_{x \in X} d \log \lambda_θ(x, x)/dθ - \int_W (d \log \lambda_θ(x, x)/dθ)λ_θ(x, x) \, dx
\]  

provides an unbiased estimating equation s(θ) = 0 (assuming in (0.26) that (d/dθ) \int_W \cdots = \int_W (d/dθ) \cdots ). This can be solved using spatstat if λ_θ is on a log linear form (Baddeley and Turner, 2000).

Second, suppose that X may have points outside W and X is Gibbs with interaction radius R. To account for edge effects we consider the conditional distribution X \( \cap \ W_\ominus R | X \cap \partial W_\ominus R = \partial x \). By (0.19) and (0.25) the log pseudo-likelihood function is then given by

\[
\sum_{x \in X \cap \partial W_\ominus R} \lambda_θ(x, x) - \int_{W_\ominus R} \lambda_θ(x, x) \, dx.
\]


0.1.4.6 Simulation-based Bayesian inference Suppose we consider a parametric Gibbs point process model with a prior on the parameter θ. Since MCMC methods are needed for simulations of the posterior distribution of θ, the main difficulty is that the Hastings ratio in a ‘conventional’ Metropolis-Hastings algorithm involves a ratio c_θ/c_θ’ of normalizing constants which we cannot compute (here θ denotes a current value of the Markov chain with invariant distribution equal to the posterior, and θ’ denotes a proposal for the next value of the chain). Heikkinen and Penttinen (1999) avoided this problem by just estimating the posterior mode, while Berthelsen and Møller (2003) estimated each ratio c_θ/c_θ’ by path sampling (Section 0.1.4.4) which at each update of the Metropolis-Hastings algorithm involved several other MCMC chains. Recently, Møller et al. (2006) introduced an auxiliary variable method which avoids such approximations. The method has been used for semi- or non-parametric Gibbs point process models (Berthelsen and Møller, 2004, 2006, 2007).
This section provides a survey of Berthelsen and Møller (2007), where the cell data in the left panel in Figure 0.10 was analyzed by a pairwise interaction process,

$$f_\theta(x) = \frac{1}{c_\theta} \prod_{x \in x} \beta(x) \prod_{\{x,y\} \subseteq x} \varphi(||x - y||),$$

(0.27)

ignoring edge effects and considering $x$ as a finite subset of the rectangular observation window in Figure 0.10; we shall later impose a very flexible prior for the first and second order interaction terms $\beta$ and $\varphi$. The data set has also been analyzed by Nielsen (2000) by transforming a Strauss point process model in order to account for inhomogeneity in the horizontal direction (Jensen and Nielsen, 2000; Nielsen and Jensen, 2004). The centre panel in Figure 0.10 clearly shows that a Poisson process is an inadequate model for the data, where the low values of the estimated pair correlation $\hat{g}(r)$ for distances $r < 0.01$ indicates repulsion between the points, so in the sequel we assume that $\varphi \leq 1$. The right panel in Figure 0.10 shows simulated 95%-envelopes under the fitted transformed Strauss point process, where $\hat{g}(r)$ is almost within the envelopes for small values of the distance $r$, suggesting that the transformed Strauss model captures the small scale inhibition in the data. Overall, $\hat{g}(r)$ follows the trend of the 95%-envelopes, but it falls outside the envelopes for some values.

We assume a priori that $\beta(x) = \beta(x_1)$ is homogeneous in the second coordinate of $x = (x_1, x_2)$, where $\beta(x_1)$ is similar to the random intensity of a Thomas process (but with $p = 1$ and using a one-dimensional Gaussian kernel in (0.4)), and where various prior assumptions on the parameters of the shot-noise process are imposed (see Berthelsen and Møller (2007)). The left panel in Figure 0.11 shows five simulated realizations of $\beta$ under its prior distribution. We also use the prior model in Berthelsen and Møller (2007) for the pairwise interaction.
function. First, \( \varphi(r) \) is assumed to be a continuous and piecewise linear function which increases from zero to one, see the right panel in Figure 0.11. The change points of \( \varphi(r) \) are modelled by a homogeneous Poisson process on \([0, r_{\text{max}}]\), where \( r_{\text{max}} = 0.02 \) (\( \hat{g} \) in Figure 0.10 indicates that there is little interaction beyond an inter-point distance of 0.01, but to be on the safe side we let \( r_{\text{max}} = 0.02 \)). The step sizes of \( \varphi(r) \) at the change points are modelled by a certain Markov chain to obtain some smoothness in \( \varphi \), and various prior assumptions on the intensity of the Poisson process of change points and the parameters of the Markov chain for the step sizes are imposed (see Berthelsen and Møller (2007)). We also assume independence between the collection of random variables \( \theta_1 \) modelling \( \beta \) and the collection of random variables \( \theta_2 \) modelling \( \varphi \). Next, since a posterior analysis indicated the need for an explicit hard core parameter \( h < r_{\text{max}} \) in the model, we modify (0.27) by replacing \( \varphi(r) \) throughout by

\[
\tilde{\varphi}(r; h) = \begin{cases} 
0 & \text{if } r < h \\
\varphi\left(\frac{(r-h)r_{\text{max}}}{r_{\text{max}}-h}\right) & \text{if } h \leq r \leq r_{\text{max}} \\
1 & \text{if } h > r_{\text{max}}.
\end{cases}
\]

Finally, \( h \) is assumed to be uniformly distributed on \([0, r_{\text{max}}]\) and independent of \((\theta_1, \theta_2)\), whereby the posterior density of \( \theta = (\theta_1, \theta_2, h) \) can be obtained.

Details on how to simulate from the posterior distribution, including how to use the auxiliary variable method from Møller et al. (2006), are given in Berthelsen and Møller (2007). The auxiliary variable method requires perfect simulations from (0.28); see Section 0.1.4.7. The left panel of Figure 0.12 shows the posterior mean \( \text{E}(\beta(x_1)|x) \) together with pointwise 95% central posterior intervals. Also the smooth estimate of the first order term obtained by (Nielsen, 2000) is shown, where the main difference compared with \( \text{E}(\beta(x_1)|x) \) is the abrupt change of \( \text{E}(\beta(x_1)|x) \) in the interval \([0.2, 0.4]\). For locations \( x = (x_1, x_2) \) near the edges of \( W \), \( \text{E}(\beta(x_1)|x) \) is ‘pulled’ towards its prior mean as a consequence of the smoothing prior. Apart from boundary effects, since \( \beta(x_1, x_2) \) only
Fig. 0.12. Posterior mean (solid line) and pointwise 95% central posterior intervals (dotted lines) for $\beta$ (left panel) and $\tilde{\phi}$ (right panel). The left panel also shows the first order term (dashed line) estimated by a fitted transformed Strauss process, and an estimate of the cell intensity (dot-dashed line).

depends on $x_1$, we may expect that the intensity $\rho_0(x_1, x_2)$ only slightly depends on $x_2$, i.e. $\rho_0(x_1, x_2) \approx \rho_0(x_1)$ where, if we let $W = [0, a] \times [0, b]$, $\rho_0(x_1, x_2)$ is given by $1/b$ times the integral of $\rho_0(x_1, x_2)$ over $x_2 \in [0, b]$. We therefore refer to $\rho_0(x_1)$ as the cell intensity, though it is more precisely the average cell intensity in $W$ at $x_1 \in [0, a]$. A non-parametric estimate of $\rho_0(x_1)$ is given by

$$\hat{\rho}(x_1) = \frac{\sum_{(y_1, y_2) \in \mathcal{X}} \phi((x_1 - y_1)/\sigma_k)/\sigma_k}{[b \times (\Phi((1 - x_1)/\sigma_k) - \Phi(x_1/\sigma_k))]}$$

which is basically the one-dimensional edge-corrected kernel estimator of Diggle (1985) with bandwidth $\sigma_k = 0.075$ (here $\phi$ and $\Phi$ denote the density and cumulative distribution function of the standard normal distribution). The left panel of Figure 0.12 also shows this estimate. The posterior mean of $\beta(x_1)$ is not unlike $\hat{\rho}(x_1)$ except that $E(\beta(x_1)|\mathbf{x})$ is higher as would be expected due to the repulsion in the likelihood. The posterior mean of $\tilde{\phi}$ is shown in the right panel of Figure 0.12 together with pointwise 95% central posterior intervals. The figure shows a distinct hard core on the interval from zero to the observed minimum inter-point distance $d$ which is a little less than 0.006, and an effective interaction range which is no more than 0.015 (the posterior distribution of $\tilde{\phi}(r)$ is concentrated close to one for $r > 0.015$). This further confirms that $r_{\text{max}}$ was chosen sufficiently large. The corner at $r = d$ of the curve showing the posterior mean of $\tilde{\phi}(r)$ is caused by that $\tilde{\phi}(r)$ is often zero for $r < d$ (since the hard core is concentrated close to $d$), while $\tilde{\phi}(r) > 0$ for $r > d$. Further posterior results, including a model checking based on the posterior predictive distribution, can be found in Berthelsen and Møller (2007).

0.1.4.7 Simulation algorithms Consider a finite point process $\mathbf{X}$ with density $f \propto h$ with respect to the unit rate Poisson process defined on a bounded region
Inference

27

$B \in \mathcal{B}$ of area $|B| > 0$, where $h$ is a ‘known’ unnormalized density. The normal-
izing constant of the density is not assumed to be known. This section reviews
algorithms for making simulated realizations of $X$.

**Birth-death algorithms** Simulation conditional on the number of points $n(X)$
can be done using a variety of Metropolis-Hastings algorithms, e.g. using a Gibbs
sampler (Ripley, 1977, 1979) or a Metropolis-within-Gibbs algorithm, where at
each iteration a single point given the remaining points is updated, see Möller
and Waegepetersen (2003). The standard algorithms (i.e. without conditioning
on $n(X)$) are discrete or continuous time algorithms of the birth-death type,
where each transition is either the addition of a new point (a birth) or the
deletion of an existing point (a death). The algorithms can easily be extended
to birth-death-move type algorithms, where e.g. in the discrete time case the
number of points is retained in a move by using a Metropolis-Hastings update
(Norman and Filinov, 1969; Geyer and Möller, 1994).

In the discrete time case, a Metropolis-Hastings birth-death algorithm may
update a current state $X_t = x$ of the Markov chain as follows. Assume that $h$
is hereditary, and define $r(x; x) = \lambda(x; x)/|B|/(n(x) + 1)$ where $\lambda$
is the Papan-
gelou conditional intensity. With probability 0.5 propose a birth, i.e. generate
a uniform point $x$ in $B$, and accept the proposal $X_{t+1} = x \cup \{x\}$ with prob-
ability $\min\{1, r(x; x)\}$. Otherwise propose a death, i.e. select a point $x \in x$
uniformly at random, and accept the proposal $X_{t+1} = x \setminus \{x\}$ with probabil-
ity $\min\{1, 1/r(x; x \setminus \{x\})\}$. As usual in a Metropolis-Hastings algorithm, if the
proposal is not accepted, $X_{t+1} = x$. This algorithm is irreducible, aperiodic,
and time reversible with invariant dens ity $f$, and convergence
of $X_t$ towards $f$ holds under weak conditions (Preston, 1977), where local st abil-
ity of $h$ implies geometrically fast convergence (Møller, 1989). It can be extended
to a coupling construction with a dominating spatial birth-death process $D_t,
t \geq 0$, which is easy to simulate (even in equilibrium) and from which we may
easily obtain the spatial birth-death process $X_t$ from above by a dependent thin-
ning. Since the coupling construction is also used below in connection to perfect
simulation we give here the details. Assume that $h$ is locally stable with uniform
upper bound $K \geq \lambda$. Initially, suppose that $X_0 \subseteq D_0$. Let $0 < \tau_1 < \tau_2 < \ldots$
denote the transition times in $D_t$, and set $\tau_0 = 0$. For $i \geq 1$, if we condition
on both $\tau_{i-1} = t$, $D_t = d$, $X_t = x$, and the history $(D_s, X_s), 0 \leq s < t$, then
$\tau_i - \tau_{i-1}$ is exponentially distributed with mean $1/(K + n(d))$. If we also condition on $\tau_i - \tau_{i-1}$, then with probability $K/(K + n(d))$ a birth occurs so that $D_{\tau_i} = d \cup \{x\}$, where $x$ is uniformly distributed on $B$, and otherwise a death happens so that both $D_{\tau_i} = d \setminus \{y\}$ and $X_{\tau_i} = x \setminus \{y\}$, where $y$ is a uniformly selected point from $d$. Furthermore, in case $D_{\tau_i} = d \cup \{x\}$, we generate a uniform number $R_i \in [0,1]$ (which is independent of $x$ and anything else so far generated), and set $X_{\tau_i} = x \cup \{x\}$ if $R_i \leq \lambda(x;x)/K$, and $X_{\tau_i} = x$ otherwise. It follows that $D_t$ is dominating $X_t$ in the sense that $X_t \subseteq D_t$ for all $t \geq 0$, and as required, $X_t$ is a spatial birth-death process with birth rate $\lambda$ and death rate 1. Moreover, $D_t$ is a spatial birth-death process with birth rate $K$ and death rate 1, so its equilibrium distribution is a homogeneous Poisson process with rate $K$.

**Perfect simulation** One of the most exciting recent developments in stochastic simulation is perfect (or exact) simulation, which turns out to be particular applicable for locally stable point processes. By this we mean an algorithm where the running time is a finite random variable and the output is a draw from a given target distribution (at least in theory — of course the use of pseudo random number generators and practical constraints of time imply that we cannot exactly return draws from the target distribution). The most famous perfect simulation algorithm is due to Propp and Wilson (1996); for a tutorial on perfect simulation in stochastic geometry, see Møller (2001).

One problem with the Propp-Wilson algorithm is that it applies only for uniformly ergodic Markov chains (Foss and Tweedie, 1998) and most spatial point process algorithms are only geometrically ergodic (Kendall, 2004). Another problem is that a kind of monotonicity property and existence of unique upper and lower bounds are needed with respect to a partial ordering, but the natural partial ordering of point processes is set inclusion, and for this ordering there is no maximal element (the minimal element is the empty point configuration $\emptyset$). Below we describe the simplest solution to these problems based on a technique called dominating coupling from the past (DCFTP). The technique was developed in Kendall (1998) and Kendall and Møller (2000); various other solutions have been proposed in Wilson (2000), Fernández et al. (2002), and Berthelsen and Møller (2002b).

DCFTP uses the coupling construction for the spatial birth-death processes $D_t$ and $X_t$ described above, assuming that $f$ is locally stable and that $D_0$ follows its Poisson process equilibrium distribution. However, we generate $D_t$ backwards in time $t \leq 0$, which by reversibility is the same stochastic construction as generating it forwards in time $t \geq 0$. One possibility, here called DCFTP1, would be to stop the backwards generation of $D_t$ at time $\tau = \sup\{t \leq 0 : D_t = \emptyset\}$, and then set $X_\tau = \emptyset$ and use the same coupling construction as before when generating forwards $X_t$, $\tau < t \leq 0$. In fact, with probability one, the algorithm terminates (i.e. $\tau > -\infty$) and $X_0 \sim f$ is a perfect simulation. Note that we do not need to generate the exponential waiting times for transitions, since only the jumps of births and deaths are used, that is, we need only to generate $J_0 = D_0$.
and if \( J_0 \neq \emptyset \), the jump chain \( J_1, J_2, \ldots, J_M \) corresponding to \( D_t \) backwards in time \( 0 > t \geq \tau \) (so that \( J_1 \neq \emptyset, \ldots, J_{M-1} \neq \emptyset \), and \( J_M = \emptyset \)) and the jump chain \( I_{-M}, I_{-M+1}, \ldots, I_0 \) corresponding to \( X_t \) forwards in time \( \tau < t \leq 0 \) (so that \( I_0 = X_0 \)). However, as demonstrated in Berthelsen and Møller (2002a), DCFTP1 is infeasible in practice, since the number of jumps \( M_1 \) can be extremely large for spatial point process models with even a modest value of \( K|B| \) (the expected number of points in \( D_0 \)).

In practice we use instead the DCFTP2-algorithm based on a sequence of discrete starting times \( T_0 = 0 > T_1 > T_2 > \ldots \) for so-called upper and lower bounding processes \( U_n^0 \) and \( L_n^0 \), \( t = T_n, T_n + 1, \ldots, 0, n = 1, 2, \ldots \), constructed forwards in time as described below such that the following funnelling/sandwiching property is satisfied,

\[
L_{t+1}^n \subseteq L_n^0 \subseteq I_t \subseteq U_n^0 \subseteq U_{t+1}^n \subseteq J_t, \quad t = T_n, T_n + 1, \ldots, 0, \quad n = 1, 2, \ldots
\]

Here we imagine for the moment that \( (D_t, X_t) \) is extended further back in time, which at least in theory is easily done, since the process regenerates each time \( D_t = \emptyset \). Thereby we see that \( L_n^0 = U_n^0 = I_0 = X_0 \) is a perfect simulation for all sufficiently large \( n \), and we let \( M_2 \) denote the smallest possible \( n \) with \( L_n^0 = U_n^0 \). The DCFTP2-algorithm consists in first generating \( D_0 \) from its Poisson process equilibrium distribution, second if \( D_0 = \emptyset \) to return \( X_0 = \emptyset \), and else for \( n = 1, \ldots, M_2 \), generating \( J_{T_n-1+1}, \ldots, J_{T_n} \) and \( (U_{T_n}^n, L_{T_n}^n), \ldots, (U_0^n, L_0^n) \), and returning \( X_0 = L_0^{M_2} \). Note that only the dominating jump chain and the pairs of upper and lower processes are generated until we obtain coalescence \( L_0^n = U_0^n \); it is not required to generate the ‘target’ jump chain \( I_t \). Empirical findings in Berthelsen and Møller (2002a) show that \( M_2 \) can be much smaller than \( M_1 \).

To specify the construction of upper and lower processes, recall that in case of a forwards birth (corresponding to a backwards death) of \( D_t \), we generated a uniform number used for determining whether a birth should occur or not in \( X_t \). Let \( R_0, R_1, \ldots \) be mutually independent uniform numbers in \([0, 1]\), which are independent of the dominating jump chain \( J_t \). To obtain (0.29) we reuse these uniform numbers for all pairs of upper and lower processes as follows. For \( n = 1, 2, \ldots \), initially, \( U_{T_n}^n = J_{T_n} \) and \( L_{T_n}^n = \emptyset \). Further, for \( t = T_n + 1, \ldots, 0 \), if \( J_t = J_{t-1} \cup \{x\} \), define

\[
\max \lambda_t^n(x) = \max \{ \lambda(x; x) : L_{t-1}^n \subseteq x \subseteq U_{t-1}^n \},
\]

\[
\min \lambda_t^n(x) = \min \{ \lambda(x; x) : L_{t-1}^n \subseteq x \subseteq U_{t-1}^n \},
\]

\[
U_t^n = \begin{cases} 
U_{t-1}^n \cup \{x\} & \text{if } R_t \leq \max \lambda_t^n(x), \\
U_{t-1}^n & \text{otherwise,}
\end{cases}
\]

\[
L_t^n = \begin{cases} 
L_{t-1}^n \cup \{x\} & \text{if } R_t \leq \min \lambda_t^n(x), \\
L_{t-1}^n & \text{otherwise.}
\end{cases}
\]
Furthermore, if $J_t = J_{t-1} \setminus \{x\}$, define $U^n_t = U_{t-1} \setminus \{x\}$ and $L^n_t = L_{t-1} \setminus \{x\}$. Hence by induction we obtain (0.29) if we are reusing the $R_i$ when constructing $I_t$ as in DCFTP1.

In practice we need some kind of monotonicity so that $\max \lambda^n_t$ and $\min \lambda^n_t$ can be quickly calculated. If $\lambda(x; x)$ is a non-decreasing function of $x$ (the ’attractive case’), i.e. $\lambda(x; y) \leq \lambda(x; x)$ whenever $y \subset x$, then $\max \lambda^n_t(x) = \lambda(x; U^n_{t-1})$ and $\min \lambda^n_t(x) = \lambda(x; L^n_{t-1})$. If instead $\lambda(x; x)$ is a non-increasing function of $x$ (the ’repulsive case’), then $\max \lambda^n_t(x) = \lambda(x; L^n_{t-1})$ and $\min \lambda^n_t(x) = \lambda(x; U^n_{t-1})$.

As starting times $T_i$, we may use a doubling scheme $T_{i+1} = 2T_i$ as proposed in Propp and Wilson (1996). In Berthelsen and Møller (2002a) a random doubling scheme is used, where $T_1 = \sup\{t \leq 0 : J_t \cap J_0 = \emptyset\}$.

For instance, DCFTP2 has been used in connection to the pairwise interaction process fitting the cell data in Section 0.1.4.6. Figure 0.8 shows perfect simulations of different Strauss processes as defined in Section 0.1.4.2, with $\theta_2 = 0$ (the Poisson case), $\theta_2 = -\log 2$, and $\theta_2 = -\infty$ (the hard core case), using the same dominating process (and associated marks $R_i$) in all three cases, with $K = \exp(\theta_1) = 100$. Due to the dependent thinning procedure in the algorithm, the point pattern with $\theta_2 = 0$ contains the two others, but the point pattern with $\theta_2 = -\infty$ does not contain the point pattern with $\theta_2 = -\log 2$ because the Strauss process is repulsive.

Acknowledgments: Much of this contribution is based on previous work with my collaborators, particularly, Adrian Baddeley, Kasper K. Berthelsen, Janine Illian, Wilfrid Kendall, and not at least Rasmus P. Waagepetersen. Supported by the Danish Natural Science Research Council, grant no. 272-06-0442 (‘Point process modelling and statistical inference’).
REFERENCES

References


Heikkinen, J. and Penttinen, A. (1999). Bayesian smoothing in the estimation of


Nielsen, L. S. and Jensen, E. B. V. (2004). Statistical inference for trans-


